The Normal or Gaussian Distribution

November 3, 2010

The Normal Distribution

The normal distribution is one of the most commonly used probability distribution for applications.

1. When we repeat an experiment numerous times and average our results, the random variable representing the average or mean tends to have a normal distribution as the number of experiments becomes large.
2. The previous fact, which is known as the central limit theorem, is fundamental to many of the statistical techniques we will discuss later.

The Normal Distribution

The normal distribution is one of the most commonly used probability distribution for applications.

1. When we repeat an experiment numerous times and average our results, the random variable representing the average or mean tends to have a normal distribution as the number of experiments becomes large.
2. The previous fact, which is known as the central limit theorem, is fundamental to many of the statistical techniques we will discuss later.
3. Many physical characteristics tend to follow a normal distribution. - Heights, weights, etc.
4. Errors in measurement or production processes can often be approximated by a normal distribution.

The Normal Distribution

The normal distribution is one of the most commonly used probability distribution for applications.

1. When we repeat an experiment numerous times and average our results, the random variable representing the average or mean tends to have a normal distribution as the number of experiments becomes large.
2. The previous fact, which is known as the central limit theorem, is fundamental to many of the statistical techniques we will discuss later.
3. Many physical characteristics tend to follow a normal distribution. - Heights, weights, etc.
4. Errors in measurement or production processes can often be approximated by a normal distribution.

Random variables with a normal distribution are said to be normal random variables.

The Normal Distribution

The normal distribution $N(\mu, \sigma)$ has two parameters associated with it:

1. The mean μ
2. The standard deviation σ.

The Normal Distribution

The normal distribution $N(\mu, \sigma)$ has two parameters associated with it:

1. The mean μ
2. The standard deviation σ.

The probability density function $f(x)$ of $N(\mu, \sigma)$ is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}.$$

The normal density function cannot be integrated in closed form. We use tables of cumulative probabilities for a special normal distribution to calculate normal probabilities.

The Normal Distribution - Properties

1. Expected Value: $E(X) = \mu$ for a normal random variable X.
2. Variance: $V(X) = \sigma^2$.

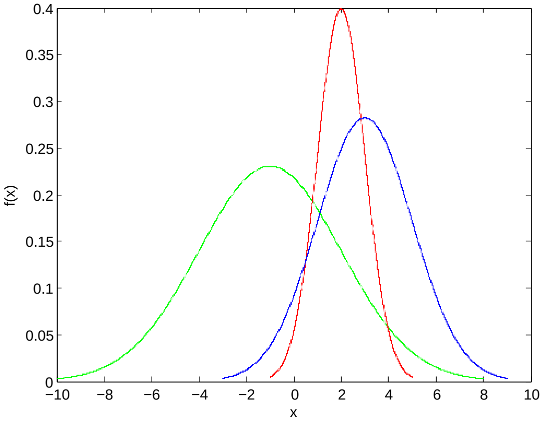
The Normal Distribution - Properties

1. Expected Value: $E(X) = \mu$ for a normal random variable X.
2. Variance: $V(X) = \sigma^2$.
3. Symmetry: The probability density function f of a normal random variable is symmetric about the mean. Formally

$$f(\mu - x) = f(\mu + x)$$

for all real x.

The Normal Distribution



The parameter μ determines the location of the distribution while σ determines the width of the bell curve.

The normal distribution with mean 0 and standard deviation 1

$$N(0, 1)$$

is called the standard normal distribution.

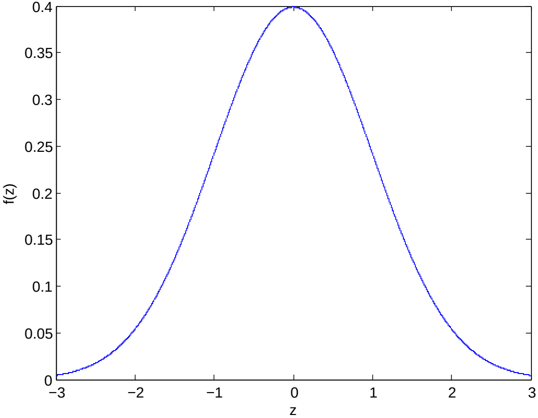
A random variable with the standard normal distribution is called a standard normal random variableand is usually denoted by Z.

The cumulative probability distribution of the standard normal distribution

$$P(Z \leq z)$$

has been tabulated and is used to calculate probabilities for any normal random variable.

The Standard Normal Distribution



The shape of the standard normal distribution is shown above.

Standard Normal Distribution

- $P(Z \leq z_0)$ gives the area under the curve to the left of z_0.
- $P(z_0 \leq Z \leq z_1) = P(Z \leq z_1) - P(Z \leq z_0)$.
- The distribution is symmetric. $P(Z \leq z_0) = P(Z \geq -z_0)$.

Example

Suppose Z is a standard normal random variable. Calculate

(i) $P(Z \leq 1.1)$;

(ii) $P(Z > 0.8)$;

(iii) $P(Z \leq -1.52)$;

(iv) $P(0.4 \leq Z \leq 1.32)$.

(v) $P(-0.2 \leq Z \leq 0.34)$.

Example

(i) $P(Z \leq 1.1)$: This can be read directly from the table. $P(Z \leq 1.1) = 0.864$.

(ii) $P(Z > 0.8) = 1 - P(Z \leq 0.8) = 1 - 0.788 = 0.212$.

(iii) $P(Z \leq -1.52)$:

Example

(i) $P(Z \leq 1.1)$: This can be read directly from the table. $P(Z \leq 1.1) = 0.864$.

(ii) $P(Z > 0.8) = 1 - P(Z \leq 0.8) = 1 - 0.788 = 0.212$.

(iii) $P(Z \leq -1.52)$: Again, we can read this directly from the table. $P(Z \leq -1.52) = 0.064$

Example

(iv) $P(0.4 \leq Z \leq 1.32)$.

Example

(iv) $P(0.4 \leq Z \leq 1.32)$. To calculate this, we note that

$$\begin{aligned} P(0.4 \leq Z \leq 1.32) &= P(Z \leq 1.32) - P(Z < 0.4) \\ &= 0.907 - 0.655 \\ &= 0.252 \end{aligned}$$

(v) Similarly,

$$P(-0.2 \leq Z \leq 0.34)$$

Example

(iv) $P(0.4 \leq Z \leq 1.32)$. To calculate this, we note that

$$\begin{aligned} P(0.4 \leq Z \leq 1.32) &= P(Z \leq 1.32) - P(Z < 0.4) \\ &= 0.907 - 0.655 \\ &= 0.252 \end{aligned}$$

(v) Similarly,

$$\begin{aligned} P(-0.2 \leq Z \leq 0.34) &= P(Z \leq 0.34) - P(Z < -0.2) \\ &= 0.633 - 0.421 \\ &= 0.212 \end{aligned}$$

Standard Normal Distribution

Example

Determine the value of z_0 such that:

(i) $P(-z_0 \leq Z \leq z_0) = 0.95$;

(ii) $P(Z \leq z_0) = 0.95$;

(iii) $P(-z_0 \leq Z \leq z_0) = 0.99$;

(iv) $P(Z \leq z_0) = 0.99$

Standard Normal Distribution

Example

(i) If $P(-z_0 \leq Z \leq z_0) = 0.95$, then $P(Z > z_0) + P(Z < -z_0) = 0.05$. By symmetry, this means that

$$P(Z > z_0) = 0.25 \text{ or } P(Z \leq z_0) = 0.975.$$

From the table of cumulative normal probabilities, the value of z_0 is 1.96

(ii) This time, we require that

$$P(Z \leq z_0) = 0.95.$$

Example

(i) If $P(-z_0 \leq Z \leq z_0) = 0.95$, then $P(Z > z_0) + P(Z < -z_0) = 0.05$. By symmetry, this means that

$$P(Z > z_0) = 0.25 \text{ or } P(Z \leq z_0) = 0.975.$$

From the table of cumulative normal probabilities, the value of z_0 is 1.96

(ii) This time, we require that

$$P(Z \leq z_0) = 0.95.$$

Using the table again, we find that the value of z_0 is 1.645.

Example

(iii) As in part (i), we are looking for a value z_0 such that

$$P(Z \leq z_0) = 0.995.$$

Example

(iii) As in part (i), we are looking for a value z_0 such that

$$P(Z \leq z_0) = 0.995.$$

From the table of normal probabilities, the value of z_0 is 2.58.

(iv) Finally, using the table, the value of z_0 for which $P(Z \leq z_0) = 0.99$ is

Example

(iii) As in part (i), we are looking for a value z_0 such that

$$P(Z \leq z_0) = 0.995.$$

From the table of normal probabilities, the value of z_0 is 2.58.

(iv) Finally, using the table, the value of z_0 for which $P(Z \leq z_0) = 0.99$ is 2.33.

Standardising

The key fact needed to calculate probabilities for a general normal random variable is the following.

Theorem

If X is a normal random variable with mean μ and standard deviation σ, then

$$Z = \frac{X - \mu}{\sigma}$$

is a standard normal random variable.

This means that to calculate $P(X \leq x)$ is the same as calculating

$$P(Z \leq \frac{x - \mu}{\sigma}).$$

Example

The actual volume of soup in 500ml jars follows a normal distribution with mean 500ml and variance 16ml. If X denotes the actual volume of soup in a jar, what is

(i) $P(X > 496)$?;

(ii) $P(X < 498)$?;

(iii) $P(492 < X < 512)$?

(iv) $P(X > 480)$?

Normal Distribution Examples

Example

(i)

$$\begin{aligned} P(X > 496) &= P(Z > \frac{496 - 500}{4}) \\ &= P(Z > -1) = 1 - 0.159 = 0.841. \end{aligned}$$

(ii)

$$P(X < 498)$$

Normal Distribution Examples

Example

(i)

$$
\begin{aligned}
P(X > 496) &= P(Z > \frac{496 - 500}{4}) \\
&= P(Z > -1) = 1 - 0.159 = 0.841.
\end{aligned}
$$

(ii)

$$
\begin{aligned}
P(X < 498) &= P(Z < \frac{498 - 500}{4}) \\
&= P(Z < -0.5) = 0.309
\end{aligned}
$$

Example

(iii)

$$P(492 < X < 506) \quad =$$

Normal Distribution Examples

Example

(iii)

$$\begin{aligned} P(492 < X < 506) &= P(\frac{492-500}{4} < Z < \frac{506-500}{4}) \\ &= P(-2 < Z < 1.5) \\ &= P(Z < 1.5) - P(Z \leq -2) \\ &= 0.933 - 0.023 = 0.91. \end{aligned}$$

(iv)

$$P(X > 493) \quad =$$

Normal Distribution Examples

Example

(iii)

$$
\begin{aligned}
P(492 < X < 506) &= P\left(\frac{492-500}{4} < Z < \frac{506-500}{4}\right) \\
&= P(-2 < Z < 1.5) \\
&= P(Z < 1.5) - P(Z \leq -2) \\
&= 0.933 - 0.023 = 0.91.
\end{aligned}
$$

(iv)

$$
\begin{aligned}
P(X > 493) &= P\left(Z > \frac{493-500}{4}\right) \\
&= P(Z > -1.75) \\
&= 1 - P(Z \leq -1.75) = 1 - 0.04 = 0.96.
\end{aligned}
$$

Normal Distribution Examples

Example

In the previous example, suppose that the mean volume of soup in a jar is unknown but that the standard deviation is 4. If only 3% of jars are to contain less than 492ml what should the mean volume of soup in a jar be?

We want the value of μ for which

Normal Distribution Examples

Example

In the previous example, suppose that the mean volume of soup in a jar is unknown but that the standard deviation is 4. If only 3% of jars are to contain less than 492ml what should the mean volume of soup in a jar be?

We want the value of μ for which

$$P(Z < \frac{492 - \mu}{4}) = 0.03.$$

From the table

Example

In the previous example, suppose that the mean volume of soup in a jar is unknown but that the standard deviation is 4. If only 3% of jars are to contain less than 492ml what should the mean volume of soup in a jar be?

We want the value of μ for which

$$P(Z < \frac{492 - \mu}{4}) = 0.03.$$

From the table

$$P(Z < -1.88) = 0.03.$$

So

Example

In the previous example, suppose that the mean volume of soup in a jar is unknown but that the standard deviation is 4. If only 3% of jars are to contain less than 492ml what should the mean volume of soup in a jar be?

We want the value of μ for which

$$P(Z < \frac{492-\mu}{4}) = 0.03.$$

From the table

$$P(Z < -1.88) = 0.03.$$

So

$$\begin{aligned} \frac{492-\mu}{4} &= -1.88 \\ \mu &= 492 + 4(1.88) = 499.52. \end{aligned}$$

Normal Approximation to the Binomial Distribution

1. The normal distribution can be used to approximate binomial probabilities when there is a very large number of trials and when both np and $n(1-p)$ are both large.
2. A rule of thumb is to use this approximation when both np and $n(1-p)$ are greater than 5. If both are greater than 15 then the approximation should be good.

Normal Approximation to the Binomial Distribution

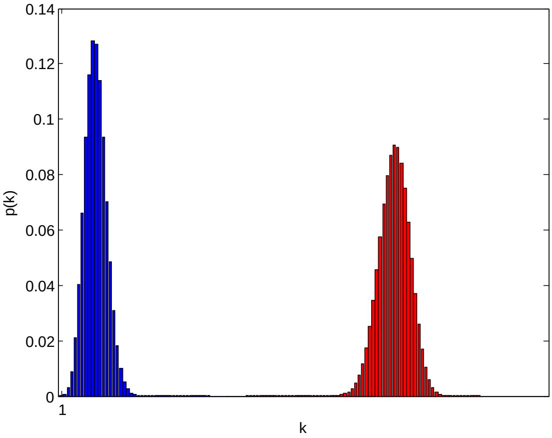
1. The normal distribution can be used to approximate binomial probabilities when there is a very large number of trials and when both np and $n(1-p)$ are both large.
2. A rule of thumb is to use this approximation when both np and $n(1-p)$ are greater than 5. If both are greater than 15 then the approximation should be good.

In this case, when X is a binomial random variable,

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

is approximately a standard normal random variable.

Normal Approximation to the Binomial Distribution



The two examples shown above are graphs of binomial probabilities with $n = 90, 120$ and $p = 0.12, 0.8$ respectively.

Normal Approximation to the Binomial Distribution

To improve the accuracy of the approximation, we usually use a correction factor to take into account that the binomial random variable is discrete while the normal is continuous.

Normal Approximation to the Binomial Distribution

To improve the accuracy of the approximation, we usually use a correction factor to take into account that the binomial random variable is discrete while the normal is continuous.

The basic idea is to treat the discrete value k as the continuous interval from $k - 0.5$ to $k + 0.5$.

Example

12% of the memory cards made at a certain factory are defective. If a sample of 150 cards is selected randomly, use the normal approximation to the binomial distribution to calculate the probability that the sample contains:

(i) at most 20 defective cards;

(ii) between 15 and 23 defective cards;

(iii) exactly 17 defective cards.

Example

(i) With the correction factor, we wish to calculate $P(X \leq 20.5)$. This is approximated by

$$\begin{aligned} P(Z \leq \frac{20.5 - 18}{\sqrt{150(.12)(.88)}}) &= P(Z \leq 0.63) \\ &= 0.736. \end{aligned}$$

Example

(i) With the correction factor, we wish to calculate $P(X \leq 20.5)$. This is approximated by

$$P(Z \leq \frac{20.5 - 18}{\sqrt{150(.12)(.88)}}) = P(Z \leq 0.63) = 0.736.$$

(ii) This time we want

$$P(\frac{14.5 - 18}{\sqrt{150(.12)(.88)}} \leq Z \leq \frac{23.5 - 18}{\sqrt{150(.12)(.88)}})$$
$$= P(-0.88 \leq Z \leq 1.38) = 0.916 - 0.189 = 0.727.$$

Example

(iii) Using the continuous correction factor, the probability we want is $P(16.5 \le X \le 17.5)$, which is

$$P(\frac{16.5-18}{\sqrt{150(.12)(.88)}} \le Z \le \frac{17.5-18}{\sqrt{150(.12)(.88)}})$$
$$= P(-0.38 \le Z \le -0.13) = 0.448 - 0.352 = 0.096$$